Blackwell–Sherman–Stein Theorem in Categorical Probability

Tomáš Gonda

In collaboration with: Tobias Fritz, Paolo Perrone, and Eigil Rischel

Applied Category Theory Conference

July 6, 2020

PERIMETER INSTITUTE

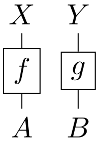
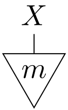
UNIVERSITY OF WATERLOO

Categorical probability

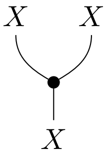
- Do measure-theory in a model agnostic way.
- Diagrammatic reasoning about probabilistic concepts.
- Synthetic proofs can apply to basic probability, stochastic processes, etc.
- Language of probabilistic programming. Aims to simplify implementation of Bayesian inference.
- Our setting: Markov categories[1]

[1] Fritz. *A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics* (2019)

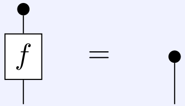
Elements of Markov categories

Diagram	Name	Interpretation
	morphism $f\colon A \to X$	A process with "random" output. Markov kernel.
	tensor product $f \otimes g$	Processes running in parallel, independently.
	state $m\colon I \to X$	A probability measure.

Elements of Markov categories

Diagram	Name	Interpretation
	$\mathrm{del}\colon X \to I$	Forget or discard X.
	$\mathrm{copy}\colon X \to X \otimes X$	Produce two copies of X.

Normalization of probabilities:

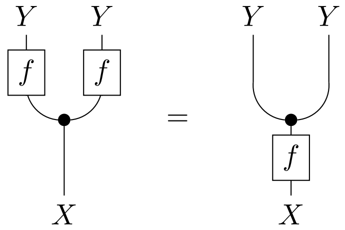


Examples of Markov categories

- *FinStoch*: Finite sets and stochastic matrices.
- *BorelStoch*: Standard Borel spaces and Markov kernels.
- *Stoch*: Measurable spaces and Markov kernels.
- *Gauss*: Vector spaces and affine maps with Gaussian noise.
- $\mathit{Fun}(I, C)$: 'Functor category' with C Markov and I small.
- Any hypergraph category (such as *Rel*).
- Any category of commutative comonoids in an SMC D.

Deterministic morphisms

f is **deterministic** if
"applying it to the same input always produces the same output".



- In *FinStoch* they are **functions**.
- In *Stoch* and *BorelStoch*, **measurable maps**.
- In *Gauss*, affine maps with **no noise**.

A toy example

- Three hypotheses: $\Theta = \{v, a, 0\}$.
  - v: Virus present.
  - a: After disease. Antibodies present.
  - 0: No encounter with the virus.

- Two tests: $V, A\colon \Theta \to X$. Two outcomes: $X = \{+, -\}$.

$V(X\|\Theta)$	v	a	0
$+$	.9	.5	.1
$-$	.1	.5	.9

$A(X\|\Theta)$	v	a	0
$+$	.76	.6	.44
$-$	.24	.4	.56

- Is one test better than the other?

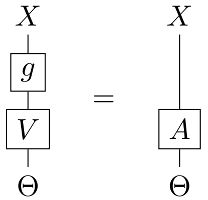
Comparison of experiments

Via simulation

$V \succeq A$ if:

V is sufficient for A, i.e.

$\exists g :$



Yes, $g = \frac{1}{5}\begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}$ works.

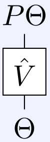
$$\frac{1}{5}\begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} .9 & .5 & .1 \\ .1 & .5 & .9 \end{pmatrix} = \\ = \begin{pmatrix} .76 & .6 & .44 \\ .24 & .4 & .56 \end{pmatrix}$$

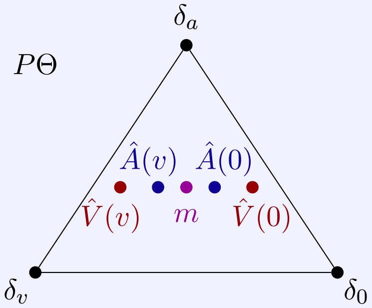
$$A(\cdot|a) = \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix} = \\ = A(\cdot|v) + A(\cdot|0)$$

$\Theta = \{v, a, 0\} \qquad X = \{+, -\}$

Comparison of experiments

For a prior m, find the posterior $\hat{V}(\theta)$ given $\theta \in \Theta$.



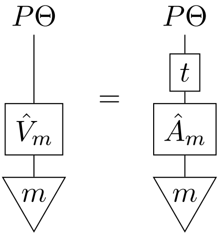


Via 2nd-order dominance

$V \succeq A$ if:

V allows one to distinguish between hypotheses better, i.e.

$\exists$ a dilation $t\colon P\Theta \to P\Theta$ with



Comparison of experiments

Via simulation

$V \succeq A$ if:

V is sufficient for A, i.e.

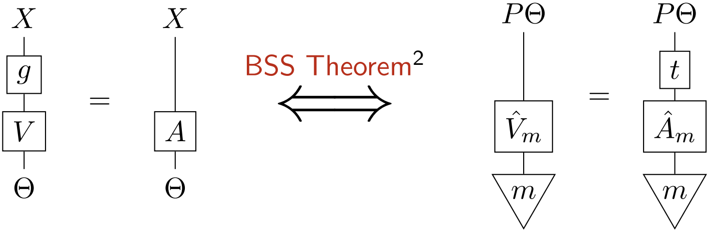
$\exists$ a kernel $g\colon X \to X$ s.t.

Via 2^nd^-order dominance

$V \succeq A$ if:

V allows one to distinguish between hypotheses better, i.e.

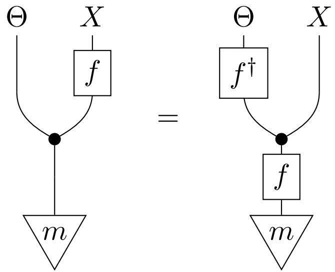
$\exists$ a dilation $t\colon P\Theta \to P\Theta$ with



[2] Blackwell. *Comparison of Experiments* (1951)

Bayesian inverses in Markov categories

$f^\dagger : X \to \Theta$ is a **Bayesian inverse** of $f : \Theta \to X$ if[3]



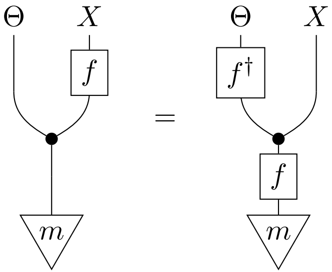
$$P(x|\theta) \cdot P(\theta) = P(\theta|x) \cdot P(x)$$

for all $\theta \in \Theta$, $x \in X$

[3] Cho, Jacobs. *Disintegration and Bayesian inversion via string diagrams* (2019)

Bayesian inverses in Markov categories

$f^\dagger \colon X \to \Theta$ is a **Bayesian inverse** of $f \colon \Theta \to X$ if[3]



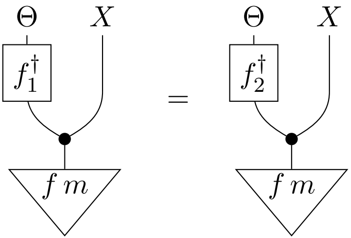
$$f(x|\theta) \cdot m(\theta) = f^\dagger(\theta|x) \cdot f\, m(x)$$

for all $\theta \in \Theta$, $x \in X$

[3] Cho, Jacobs. *Disintegration and Bayesian inversion via string diagrams* (2019)

Almost surely

If $f_1^\dagger$ and $f_2^\dagger$ are Bayesian inverses of f, then



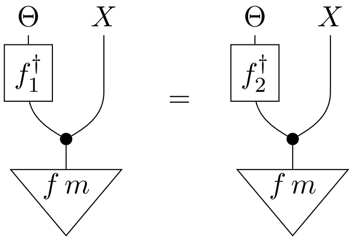
They are equal, $f\,m$-almost surely, in *FinStoch*:

$$f_1^\dagger(\theta|x) = f_2^\dagger(\theta|x)$$

for all θ and those x with $f\,m(x) > 0$.

Almost surely

If $f_1^\dagger$ and $f_2^\dagger$ are Bayesian inverses of f, then

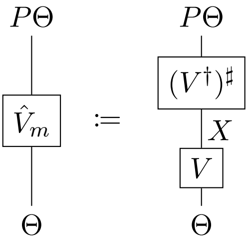


They are equal, $f\,m$-almost surely, denoted

$$f_1^\dagger =_{f m\text{-a.s.}} f_2^\dagger$$

Standard experiment

- The **standard experiment of V** is



where $(V^\dagger)^\sharp$ maps x to $V^\dagger(x)$.

- "$(V^\dagger)^\sharp$ is the deterministic analogue of $V^\dagger$"

Distribution objects

- We would like:

  stochastic matrices $X \to Y \cong$ functions $X \to PY$

- In *BorelStoch*,

  Markov kernels $X \to Y \cong$ measurable maps $X \to PY$

- In a **representable Markov category** C,

$$C(X, Y) \cong C_{\text{det}}(X, PY)$$

- Then $P\colon C \to C_{\text{det}}$ is right adjoint of $C_{\text{det}} \hookrightarrow C$.
- In *BorelStoch*, P gives the Giry monad.

Unit and counit

In a representable Markov category C,

$$\mathsf{C}(X,Y) \cong \mathsf{C}_{\mathrm{det}}(X,PY)$$

$$\mathrm{id} \in \mathsf{C}(Y,Y) \xleftrightarrow{\simeq} \boldsymbol{\delta} \in \mathsf{C}_{\mathrm{det}}(Y,PY)$$

"δ maps y to the delta measure δ_y"

$$\mathbf{samp} \in \mathsf{C}(PY,Y) \xleftrightarrow{\simeq} \mathrm{id} \in \mathsf{C}_{\mathrm{det}}(PY,PY)$$

"takes a measure m and produces an m-distributed sample"

Dilations

A dilation $t\colon P\Theta \to P\Theta$ preserves barycenters.

$$\begin{array}{ccc} P\Theta & \xrightarrow{\;t\;} & P\Theta \\ {\scriptstyle \mathrm{samp}}\downarrow & & \downarrow{\scriptstyle \mathrm{samp}} \\ \Theta & = & \Theta \end{array}$$

If it commutes μ-a.s., then t is a μ-dilation.

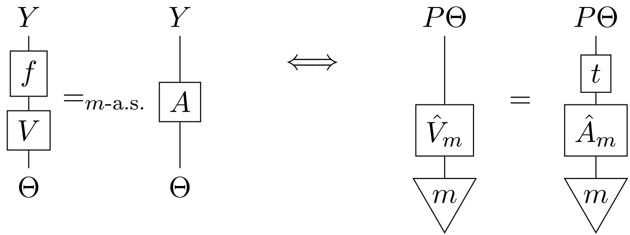
Synthetic BSS Theorem

If a Markov category

- is representable, and
- has Bayesian inverses,

then for any $m\colon I \to \Theta$

$\exists$ a morphism $f\colon X \to Y$ $\exists$ an $(\hat{A}\, m)$-dilation $t\colon P\Theta \to P\Theta$



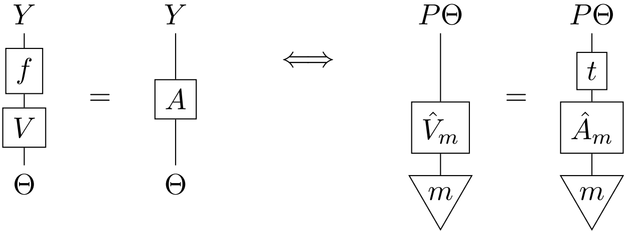
Synthetic BSS Theorem

If a Markov category

- is representable, and
- has Bayesian inverses,

then for any $m\colon I \to \Theta$ such that m-a.s. equality implies equality

$\exists$ a morphism $f\colon X \to Y$ $\qquad$ $\exists$ an $(\hat{A}\,m)$-dilation $t\colon P\Theta \to P\Theta$



Conclusions

Introduced in this talk:

- Synthetic BSS theorem for arbitrary Θ.
- Distribution objects.
- Representable Markov categories.

To look for in the full article:

- 2^{nd}-order stochastic dominance vs. partial evaluations.
- When is $\mathsf{C} \cong \mathsf{Kl}(P)$?
- Compatibility of P and a.s. equality.

Outlook:

- Abstract resource theory of distinguishability / comparison of 'structured' experiments.

Conclusions

Thank you for listening!

Main references:

- Fritz. *A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics* (2019)
- Blackwell. *Comparison of Experiments* (1951)
- Cho, Jacobs. *Disintegration and Bayesian inversion via string diagrams* (2019)